

# Optimality of experimental designs

Ryuei Nishii

*Faculty of Integrated Arts and Sciences, Hiroshima University, Hiroshima 730, Japan*

Received 4 November 1988

Revised 13 June 1989

## Abstract

Nishii, R., Optimality of experimental designs, Discrete Mathematics 116 (1993) 209–225.

In this paper, we give a survey of optimality of experimental designs. The equivalence theory between optimalities is discussed using the directional derivative of the criterion function. Optimality based on the mean squared error is also treated and, in particular, is applied to polynomial regression.

## 1. Introduction

In the theory of optimal designs the aim is to find *good* experimental designs. Here *goodness* of a design is evaluated by some real-valued function which assesses the information gained through the design. Such a function is called a *criterion function*, and is required to meet intuitive conditions (see next section for details). A typical example of such criterion functions is the determinant of the information matrix of the design. The notation used here is similar to that of Silvey [17].

Let  $\mathcal{X}$  be a design space which is a compact experimental region in the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ . Let  $\mathbf{f}(x) = (f_1(x), \dots, f_k(x))'$  be a vector of  $k$  given continuous functions of an explanatory variable  $x \in \mathcal{X}$ . We express the relationship between a response variable  $y$  and its explanatory variable  $x \in \mathcal{X}$  as follows:

$$y = \mathbf{f}(x)' \boldsymbol{\beta} + e, \quad (1.1)$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$  is a vector of unknown parameters and  $e$  is an error random variable with mean zero and variance  $\sigma^2$ . Given  $n$  independent observations  $y_1, \dots, y_n$ , based on  $n$  explanatory variables  $x_1, \dots, x_n$ , let  $X = [\mathbf{f}(x_1), \dots, \mathbf{f}(x_n)]'$ , a matrix of size  $n \times k$ , and let  $\mathbf{y} = (y_1, \dots, y_n)'$ . Representing linear relations in the matrix form, we get

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e} \quad \text{with } E(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad \text{Var}(\mathbf{e}) = \sigma^2 I_n, \quad (1.2)$$

*Correspondence to:* Ryuei Nishii, Faculty of Integrated Arts and Sciences, Hiroshima University, Hiroshima 730, Japan.

where  $\text{Var}$  stands for variance-covariance matrix. In the linear model (1.2), we apply the method of least squares to minimize the sum of squares

$$\sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (1.3)$$

Differentiating (1.3) with respect to  $\boldsymbol{\beta}$  we obtain the normal equation

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}. \quad (1.4)$$

If  $\text{rank } \mathbf{X} = k$ , then  $\mathbf{X}'\mathbf{X}$  is positive definite, and hence (1.4) has a unique solution which is

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_k)' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

This is the least squares estimator of  $\boldsymbol{\beta}$ . It is well known that

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad \text{and} \quad \text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2. \quad (1.5)$$

Furthermore,  $\hat{\boldsymbol{\beta}}$  has the least variance-covariance matrix (in the sense of the Loewner ordering) among all linear unbiased estimators of  $\boldsymbol{\beta}$ . Hence  $\hat{\boldsymbol{\beta}}$  is called the *best linear unbiased estimator* of  $\boldsymbol{\beta}$ .

The  $k \times k$  matrix  $\mathbf{X}'\mathbf{X}$  is called the *information matrix* since  $\sigma^{-2}\mathbf{X}'\mathbf{X}$  is Fisher's information with respect to the parameter  $\boldsymbol{\beta}$  when  $\mathbf{e}$  is normal. By (1.5) the accuracy of the estimate  $\hat{\boldsymbol{\beta}}$  is characterized by  $\sigma^{-2}\mathbf{X}'\mathbf{X}$  or simply  $\mathbf{X}'\mathbf{X}$ , because  $\sigma^2$  is unknown and is beyond our control. Therefore the *goodness* of a design is measured by the magnitude of the information matrix.

The explanatory variables  $x_1, \dots, x_n$  may not be distinct. Let  $x_{(1)}, x_{(2)}, \dots$  be distinct explanatory variables replicated respectively  $r_1, r_2, \dots$  times with  $\sum_i r_i = n$ , and let  $\eta_n$  be a discrete probability measure on the design space  $\mathcal{X}$  assigning probability  $r_i/n$  to  $x_{(i)}$ . Define

$$M(\eta_n) = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}(x)' d\eta_n(x) = \sum_i (r_i/n) \mathbf{f}(x_{(i)})\mathbf{f}(x_{(i)})'.$$

Then

$$\mathbf{X}'\mathbf{X} = \sum_i r_i \mathbf{f}(x_{(i)})\mathbf{f}(x_{(i)})' = n \sum_i (r_i/n) \mathbf{f}(x_{(i)})\mathbf{f}(x_{(i)})' = nM(\eta_n).$$

Hence the information matrix of an  $n$ -point design can be expressed in the form  $nM(\eta_n)$ . In this sense the probability measure  $\eta_n$  is called a *normalized design*, and  $M(\eta_n)$  is called the *information matrix of  $\eta_n$* . In the sequel we aim to find the information matrix, which is best in some sense, among the set of information matrices

$$\mathcal{M} = \{M(\eta): \eta \text{ is a probability measure on } \mathcal{X}\}, \quad (1.6)$$

where  $M(\eta) = \int_{\mathcal{X}} f(x)f(x)' d\eta(x)$  is a positive semi-definite matrix of order  $k$ . Obviously  $\mathcal{M}$  is a convex subset of the  $k(k+1)/2$ -dimensional Euclidean space. Hence Carathéodory's Theorem (see e.g., [17, Appendix 2]) shows that any information matrix in  $\mathcal{M}$  is expressible as a discrete design which has at most  $k(k+1)/2 + 1$  support points on the design space  $\mathcal{X}$ . The set  $\mathcal{M}$  is, moreover, compact because of the assumption that  $f$  is continuous and  $\mathcal{X}$  is compact. Let

$$\mathcal{M}^+ = \{M(\eta) \in \mathcal{M} : M(\eta) \text{ is positive definite}\}.$$

We assume that  $\mathcal{M}^+$  is not empty.

Examples of optimality criteria for designs are given in Section 2. Section 3 summarizes the equivalence theory. Partial optimality criteria are reviewed in Section 4. Section 5 concerns  $A$ -optimality, and Section 6 treats optimality based on mean squared error. Section 7 shows examples of optimal design and Section 8 presents open problems in polynomial regression.

The readers who are interested in this field should refer to Pukelsheim [13, 14], which gives a unified treatment of the theory of optimal designs. Atkinson [1] gives an excellent survey with emphasis on nonstandard applications of optimal design theory and provides a list of more than 170 references. Paterson [11] and Matthews [8] review optimal block designs and crossover designs respectively. Related literature can be found in [2, 5, 12].

## 2. The global optimality criteria

Optimality criteria are classified into two groups: (1) global optimality criteria and (2) partial optimality criteria. Global criteria are used when all parameters  $\beta_1, \dots, \beta_k$  are important, and partial criteria are used when only partial information on  $\beta_1, \dots, \beta_k$  is needed. In this and next sections we consider the global optimality criteria. We write  $M > 0$  when  $M$  is positive definite, and  $M \geq 0$  when  $M$  is positive semi-definite. Also  $M > N$  is used when  $M - N > 0$ .

Let a function  $\Phi$  be a measure assessing information gained from designs. Assume that  $\Phi: \mathcal{M} \mapsto \mathbb{R} \cup \{-\infty\}$  is bounded from above. The design  $\eta^*$  which maximizes  $\Phi(M(\eta))$  is called the  $\Phi$ -optimal design and  $\Phi$  is called the criterion function.  $M(\eta^*)$  is called the  $\Phi$ -optimal information matrix. For examples of  $\Phi$ , see [12, Chapter IV] and [17, pp. 10–14]. However note that the criterion functions appearing in these books evaluate loss of information. In such a case the  $\Phi$ -optimal design is defined when it minimizes  $\Phi(M(\eta))$  among all designs  $\eta$ .

When all parameters are important, designs with singular information matrices may be useless. Hence the basic requirements for the global optimality criterion function  $\Phi$  are:

- (A)  $\Phi(M) > \Phi(N)$  for any  $M > 0$  and any singular  $N$ ,
- (B)  $\Phi$  is strictly increasing, i.e.,  $\Phi(M) \geq \Phi(N)$  if  $M \geq N$ , and  $\Phi(M) > \Phi(N)$  if  $M > N$ .

The following widely used criterion functions satisfy these requirements.

(1) *The D-optimality criterion*

The *D*-optimality criterion is defined by the function

$$\Phi_D(M) = (\det M)^{1/k}$$

on  $\mathcal{M}$ . When  $e$  is normal, a confidence region of  $\beta$  is given by the  $k$ -dimensional ellipsoid

$$\{\beta : (\beta - \hat{\beta})' M (\beta - \hat{\beta}) \leq \text{constant}\},$$

where the constant is determined by the  $F$  distribution. The volume of the region is known to be proportional to  $(\det M)^{-1/2}$ . Hence a *D*-optimal design minimizes the volume of the confidence ellipsoid whatever the value of the constant is.  $\det((\sigma^2/n)M^{-1}) = \det(\text{Var}(\hat{\beta}))$  is called the *generalized variance*. Pukelsheim [14] characterizes *D*-optimality by the group acting on the model parameters. See also [18].

Note that maximizing  $\det M$  is equivalent to maximizing  $(\det M)^{1/k}$  or  $\log(\det M)$ . However  $(\det M)^{1/k}$  and  $\log(\det M)$  are concave functions, whereas the function  $\det M$  is not concave.

(2) *The A-optimality criterion*

The *A*-optimality criterion is defined by the function

$$\Phi_A(M) = \begin{cases} \left( \frac{1}{k} \text{tr } M^{-1} \right)^{-1} & \text{if } M \in \mathcal{M}^+, \\ 0 & \text{if } M \in \mathcal{M} - \mathcal{M}^+ \end{cases}$$

on  $\mathcal{M}$ . Since  $\sum_{i=1}^k \text{var}(\hat{\beta}_i) = (\sigma^2/n) \text{tr } M^{-1}$ , an *A*-optimal design suppresses the total variances of  $\hat{\beta}_1, \dots, \hat{\beta}_k$  without taking into account the correlations among these estimates. This criterion is useful when all parameters  $\beta_1, \dots, \beta_k$  are equally important.

(3) *The G-optimality criterion*

The *G*-optimality criterion is defined by the function

$$\Phi_G(M) = \begin{cases} \left( \max_{x \in \mathcal{X}} f(x)' M^{-1} f(x) \right)^{-1} & \text{if } M \in \mathcal{M}^+, \\ 0 & \text{if } M \in \mathcal{M} - \mathcal{M}^+ \end{cases}$$

on  $\mathcal{M}$ . Since  $\text{var}(\hat{\beta}' f(x)) = (\sigma^2/n) f(x)' M^{-1} f(x)$ , a  $G$ -optimal design will provide a good predictor for all linear combinations of  $\beta' f(x)$  for  $x \in \mathcal{X}$ .

(4) *The E-optimality criterion*

The  $E$ -optimality criterion is defined by the function

$$\Phi_E(M) = \text{the minimum eigenvalue of } M$$

on  $\mathcal{M}$ . An  $E$ -optimal design minimizes the maximum variance of  $c' \hat{\beta}$  with  $c'c = 1$ .

These criterion functions satisfy the basic requirements (A) and (B). Furthermore, they have the common properties: (C) continuity, (D) nonnegativity, (E) positive homogeneity and (F) concavity, i.e.,

(C)  $\Phi(M)$  is continuous if  $M \in \mathcal{M}^+$ ,

(D)  $\Phi(M) > 0$  if  $M \in \mathcal{M}^+$ ;  $\Phi(M) = 0$  if  $M \in \mathcal{M} - \mathcal{M}^+$ ,

(E)  $\Phi(cM) = c\Phi(M)$  for all positive constants  $c$ ,

(F)  $\Phi(\lambda M + \bar{\lambda} N) \geq \lambda \Phi(M) + \bar{\lambda} \Phi(N)$  for  $M, N \in \mathcal{M}$ ,  $\bar{\lambda} = 1 - \lambda$ ,  $0 < \lambda < 1$ .

Properties (D), (E) and (F) yield (B) because

$$\Phi(M) = 2\Phi\left\{\frac{1}{2}N + \frac{1}{2}(M - N)\right\} \geq \Phi(N) + \Phi(M - N) \geq \Phi(N) \quad \text{if } M \geq N,$$

and  $\Phi(M) > \Phi(N)$  if  $M > N > 0$ . Concavity (F) indicates that a mixture of designs gives more information than a simpler design gives. Furthermore, concavity is useful in considering the directional derivative of  $\Phi$ , as will be shown in Section 3.

Let us summarize the relation between the properties of  $\Phi$  and the  $\Phi$ -optimal information matrix. Continuity of  $\Phi$  assures the existence of the  $\Phi$ -optimal information matrix in  $\mathcal{M}$  because  $\mathcal{M}$  is compact. The conditions (D) and (E) of  $\Phi$  imply that its maximum occurs at the boundary points of  $\mathcal{M}$  which are positive definite. Concavity (F) implies that all  $\Phi$ -optimal information matrices constitute a convex set in  $\partial\mathcal{M} \cap \mathcal{M}^+$ , where  $\partial\mathcal{M}$  denotes a set of boundary points of  $\mathcal{M}$ .

We say that  $\Phi$  is *strictly concave* if

$$\Phi(\lambda M + \bar{\lambda} N) > \lambda \Phi(M) + \bar{\lambda} \Phi(N)$$

for  $\bar{\lambda} = 1 - \lambda$ ,  $0 < \lambda < 1$ ,  $M, N \in \mathcal{M}^+$ ,  $M$  and  $N$  not proportional. It is easy to show that  $\Phi_D$  and  $\Phi_A$  are strictly concave. If  $\Phi$  is strictly increasing and strictly concave, then the  $\Phi$ -optimal information matrix is a unique element of  $\partial\mathcal{M} \cap \mathcal{M}^+$  since distinct boundary points of  $\mathcal{M}$  are not proportional to each other. However the  $\Phi$ -optimal design is not necessarily unique since different designs may have the same information matrix.

### 3. Directional derivatives of a criterion function

Let  $\Phi$  be a criterion function which is concave. Then the general equivalence theory based on the derivative in the Fréchet sense is a useful technique for constructing the  $\Phi$ -optimal design.

By the concavity we know that the function

$$[\Phi\{(1-\varepsilon)M + \varepsilon N\} - \Phi(M)]/\varepsilon \quad \text{in } 0 < \varepsilon \leq 1$$

is non-increasing, where  $M, N \in \mathcal{M}$ . Hence the limit

$$F_\Phi(M, N) \equiv \lim_{\varepsilon \rightarrow +0} \frac{1}{\varepsilon} [\Phi\{(1-\varepsilon)M + \varepsilon N\} - \Phi(M)] \quad (\geq \Phi(N) - \Phi(M)) \quad (3.1)$$

exists when the value  $\infty$  is allowed. This is called the *Fréchet derivative of  $\Phi$  at  $M \in \mathcal{M}$  in the direction of  $N \in \mathcal{M}$* . We say that  $\Phi$  is *differentiable at  $M$*  if  $F_\Phi(M, \sum \lambda_i N_i) = \sum \lambda_i F_\Phi(M, N_i)$  for  $\sum \lambda_i = 1$  and  $\lambda_i > 0$ . The following theorem can be found in [17, pp. 19–22].

**Theorem 3.1.** *Let the criterion function  $\Phi$  be concave on  $\mathcal{M}$  and denote its directional derivative by (3.1). Then:*

(i)  $M^* = M(\eta^*)$  is  $\Phi$ -optimal if and only if for any  $M \in \mathcal{M}$ ,

$$F_\Phi(M^*, M) \leq 0.$$

(ii) Suppose  $\Phi$  is differentiable at  $M^* = M(\eta^*)$ , then  $M^*$  is  $\Phi$ -optimal on  $\mathcal{M}$  if and only if

$$\max_{x \in \mathcal{X}} F_\Phi(M^*, f(x)) f(x)' = 0.$$

In this case all support points  $x^*$  of the optimal design  $\eta^*$  satisfy

$$F_\Phi(M^*, f(x^*)) f(x^*)' = 0.$$

(iii) If  $\Phi$  is differentiable at all  $M \in \mathcal{M}^+$ , then  $M^* \in \mathcal{M}^+$  is  $\Phi$ -optimal if and only if

$$\max_{x \in \mathcal{X}} F_\Phi(M^*, f(x)) f(x)' = \min_{M \in \mathcal{M}^+} \max_{x \in \mathcal{X}} F_\Phi(M, f(x)) f(x)'.$$

This theorem shows that maximizing  $\Phi$  on  $\mathcal{M}$  is equivalent to minimizing

$$\max_{x \in \mathcal{X}} F_\Phi(M, f(x)) f(x)'$$

with respect to  $M \in \mathcal{M}$ . Henceforth we will refer to this as the *equivalence theorem*.

**Example.** Let  $\Phi(M) = \log(\det M)$ . Then  $\Phi$  is concave and differentiable at any positive definite matrix because  $F_\Phi(M, N) = \text{tr}(M^{-1}N) - k$  if  $M \in \mathcal{M}^+$ . By (iii) of Theorem 3.1 we have

$$M^* \text{ is } D\text{-optimal} \Leftrightarrow M^* \text{ minimizes } \max_{x \in \mathcal{X}} \{f(x)' M^{*-1} f(x) - k\}$$

$$\Leftrightarrow M^* \text{ is } G\text{-optimal}.$$

This is the equivalence theorem described in [6].

By the equivalence theorem, we can easily check whether a given information matrix is optimal or not. The theorem is also useful in constructing the optimal design by the iterative procedure, since the directional derivative is available for finding a direction which makes the criterion function large.

#### 4. The partial optimality criteria

In block designs, the treatment effects are usually parameters of interest and the block effects are treated as nuisance parameters. The partial optimality criteria are proposed for such designs, see Gaffke [3] and Pázman [12, Chapter IV]. In general, suppose we are expected to obtain the optimal design for estimating linear combinations of  $\beta$ , say  $K'\beta$ , where  $K$  is a given  $k \times s$  matrix of rank  $s$ . It is known that  $K'\beta$  is estimable by the design  $\xi$  if and only if the column space of  $K$  is included in that of the information matrix  $M(\xi)$ . The collection of such information matrices of  $\mathcal{M}$  is denoted by  $\mathcal{M}(K)$ . Obviously  $\mathcal{M}^+ \subset \mathcal{M}(K) \subset \mathcal{M}$ . Let  $M(\xi)^-$  be the generalized inverse of  $M(\xi)$ .  $K'\beta$  is estimable by  $\xi$  such that  $M(\xi) \in \mathcal{M}(K)$ . Then the best linear unbiased estimator of  $K'\beta$ , based on the design  $\xi$ , has the  $s \times s$  positive definite dispersion matrix  $\text{Var}(K'\hat{\beta}) = \sigma^2 K' M(\xi)^- K$ . Hence optimality is defined by the function of positive definite matrices  $K' M(\xi)^- K$  for  $M(\xi) \in \mathcal{M}(K)$ , see, for instance, [17, pp. 25–26]. Kiefer [7] has defined the  $L_p$ -class ( $p > 0$ ) of optimality criteria

$$\Phi_{K,p}(M) = \begin{cases} \{s^{-1} \text{tr}(K' M^- K)\}^{-1/p} & \text{if } M \in \mathcal{M}(K), \\ 0 & \text{if } M \in \mathcal{M} - \mathcal{M}(K). \end{cases}$$

Many optimality criteria are members of the  $L_p$ -class. For example,  $A$ -,  $D$ - and  $E$ -optimalities are given by

$$\Phi_{K,1}(M) = \{s^{-1} \text{tr}(K' M^- K)\}^{-1},$$

$$\lim_{p \rightarrow +0} \Phi_{K,p}(M) = s^{-1} \log \det(K' M^- K), \text{ and}$$

$$\lim_{p \rightarrow \infty} \Phi_{K,p}(M) = \text{the maximum eigenvalue of } K' M^- K,$$

respectively.

When  $\text{rank } K = s < k$ , the  $\Phi_{K,p}$ -optimal information matrix in  $\mathcal{M}(K)$  may be singular. In many cases, differentiability at singular matrices does not hold. Hence Theorem 3.1 cannot be applied. A similar equivalence theory for singular optimal information matrices is derived in [9]. Pukelsheim and Titterton [15] discuss general equivalence by using Lagrangian theory, which is valid for singular optimal information matrices.

### 5. $A$ -optimal designs for single polynomial regression

We now consider polynomial regression with a single variable. Let the design space be a compact interval  $[-1, 1]$  and  $f(x) = (1, x, \dots, x^{k-1})'$ . The linear model is expressed as a polynomial of order  $k-1$

$$y = \beta_0 + \beta_1 x + \dots + \beta_{k-1} x^{k-1} + e.$$

Then a probability measure  $\xi$  on  $[-1, 1]$  has an information matrix  $M$  whose  $(i, j)$ -elements are given by the  $(i+j-2)$ -th moments of  $\xi$ , i.e.,  $M = (\int_{-1}^1 x^{i+j-2} d\xi(x))$  of order  $k$ . Let  $\bar{\xi}$  be a mirror image of  $\xi$  with respect to zero. We say a probability measure  $\xi$  on  $[-1, 1]$  is *symmetric* if  $\xi = \bar{\xi}$ . In this case all odd moments of  $\xi$  are zero.

In this polynomial setting, it is well known that the  $D$ -optimal design assigns probability  $1/k$  to all roots of  $(1-x^2) \cdot P'_{k-1}(x) = 0$ , where  $P_m(x)$  is the  $m$ -th Legendre polynomial. See, e.g., [2, pp. 88–94]. In this section we consider the  $A$ -optimal design. Let  $L$  be a positive definite matrix of order  $k$  whose  $(i, j)$ -elements are zero if  $i+j$  are odd. (An example of  $L$  is the identity matrix  $I_k$ .) We define a criterion function

$$\Psi(M) = \begin{cases} -\text{tr}(M^{-1}L) & \text{if } M \in \mathcal{M}^+, \\ -\infty & \text{if } M \in \mathcal{M} - \mathcal{M}^+. \end{cases} \quad (5.1)$$

Note that  $\Psi(M(\xi)) = \Psi(M(\bar{\xi}))$  where  $\bar{\xi}$  is a mirror image of the design  $\xi$ . Note also that  $\Psi$  does not meet conditions (D) and (E) in Section 2, but  $\Psi$  ensures the conditions required in Theorem 3.1.

**Theorem 5.1.** *Let  $\Psi$  be a function defined by (5.1) with a given positive definite matrix  $L$ . Then the  $\Psi$ -optimal design  $\xi_k^*$  exists uniquely and is a discrete design assigning probability  $\lambda_i > 0$  to  $k$  support points  $-1 = x_1 < x_2 < \dots < x_k = 1$*

$$\left( \text{write } \xi_k^* = \begin{pmatrix} x_1, x_2, \dots, x_k \\ \lambda_1, \lambda_2, \dots, \lambda_k \end{pmatrix} \right).$$

*It is also symmetric, i.e.,  $x_i$  and  $\lambda_i$  satisfy*

$$x_i = -x_{k-i+1} \quad \text{and} \quad \lambda_i = \lambda_{k-i+1} \quad \text{for } i = 1, \dots, k,$$

*where  $\sum_{i=1}^k \lambda_i = 1$ .*

**Proof.** The function  $\Psi$  is continuous and strictly increasing. Hence the  $\Psi$ -optimal design (say,  $\xi_k^*$ ) exists uniquely. Let  $\mathbf{x} = (1, x, \dots, x^{k-1})'$  and

$$g(x) = F_\Psi(M(\xi_k^*), \mathbf{x}\mathbf{x}') = \mathbf{x}' M(\xi_k^*)^{-1} L M(\xi_k^*)^{-1} \mathbf{x} - \text{tr}(M(\xi_k^*)^{-1} L).$$

Then  $g(x)$  ( $-1 \leq x \leq 1$ ) is a polynomial of order  $2(k-1)$ . By Theorem 3.1,  $g(x) \leq 0$  for  $-1 \leq x \leq 1$  and  $g(x) = 0$  when  $x$  is a support point of  $\xi_k^*$ . Assume there exist more than  $k+1$  points which are the roots of  $g(x) = 0$ . This implies that the order of  $g(x)$  must not be less than  $1 + 2(k-1) + 1 = 2k$ , which contradicts with the fact that the order of  $g(x)$



is  $2(k-1)$ . The symmetry and the uniqueness of  $\xi^*$  follow from concavity and  $\Psi(M(\xi)) = \Psi(M(\bar{\xi}))$ , where  $\bar{\xi}$  is a mirror image.  $\square$

Take  $L = I_k$  in (5.1). We will obtain the  $\Psi$ -optimal design or in other words the  $A$ -optimal design. By Theorem 5.1, we can express the  $A$ -optimal design  $\xi_k^*$  as

$$\xi_k^* = \begin{pmatrix} x_1, x_2, \dots, x_k \\ \lambda_1, \lambda_2, \dots, \lambda_k \end{pmatrix}.$$

Hence

$$M(\xi_k^*) = XAX'$$

where  $X = [x_1, \dots, x_k]$  of order  $k$ ,  $x_i = (1, x_i, \dots, x_i^{k-1})'$  and  $A = \text{diag}(\lambda_1, \dots, \lambda_k)$ . Thus

$$\text{tr } M(\xi_k^*)^{-1} = \sum_{i=1}^k \Delta_{ii} / \{\lambda_i \det(X'X)\},$$

where  $\Delta_{ii}$  are  $(i, i)$ -cofactors of the matrix  $X'X$ . These terms are evaluated as

$$\det(X'X) = \Delta_k(x_1, \dots, x_k) \quad \text{and} \quad \Delta_{ii} = H(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k),$$

where

$$\Delta_t(z_1, \dots, z_t) = \prod_{1 \leq u < v \leq t} (z_u - z_v)^2 \quad \text{for } t = k-1, k$$

and

$$H(z_1, \dots, z_t) = \Delta_t(z_1, \dots, z_t) \{1 + (z_1 + \dots + z_t)^2 + (z_1 z_2 + z_1 z_3 + \dots + z_{t-1} z_t)^2 + \dots + (z_1 z_2 \dots z_t)^2\} \quad \text{for } t = k-1.$$

We obtain the next result using Lagrangian theory.

**Lemma 5.1.** *Let  $c_i = \Delta_{ii} / \det(X'X)$  for  $i = 1, \dots, k$ . Then the  $A$ -optimal design  $\xi_k^*$  satisfies*

$$\text{tr } M(\xi_k^*)^{-1} = \left( \sum_{i=1}^k \sqrt{c_i} \right)^2, \quad \text{and} \quad \lambda_i = \sqrt{c_i} / \sum_{j=1}^k \sqrt{c_j}.$$

Note that  $c_i = c_{k-i+1}$  by the symmetry of  $x_1, \dots, x_k$ . Thus we must only determine the  $l-1$  support points of  $\xi_k^*$  if  $k = 2l$  or  $2l+1$ .

When  $k = 2$  or  $3$ ,  $A$ -optimal designs are well known, namely

$$\xi_2^* = \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{and} \quad \xi_3^* = \begin{pmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}.$$

When  $k = 4$ , let

$$\xi_4^* = \begin{pmatrix} -1 & -\alpha & \alpha & 1 \\ \lambda_1 & \lambda_2 & \lambda_2 & \lambda_1 \end{pmatrix},$$

where  $\lambda_1$  and  $\lambda_2$  are functions of  $\alpha$  ( $0 < \alpha < 1$ ). Then

$$\text{tr } M(\zeta_4^*)^{-1} = 2 \left( \frac{(1+\alpha^4)^{1/2}}{1-\alpha^2} + \frac{(1+\alpha^2)^{1/2}}{\alpha(1-\alpha^2)} \right)^2.$$

Minimizing this expression with respect to  $\alpha$  ( $0 < \alpha < 1$ ), we obtain

$$\alpha = \left\{ \frac{\sqrt{7}-2}{3} \right\}^{1/2} = 0.46400,$$

$$\lambda_1 = \frac{4-\sqrt{7}}{9} = 0.15047 \quad \text{and} \quad \lambda_2 = \frac{1+2\sqrt{7}}{18} = 0.34953.$$

When  $k$  is greater than 4, it is hard to derive the exact solutions of the support points. Numerical solutions are as follows:

$$\zeta_5^* = \begin{pmatrix} -1 & -\alpha & 0 & \alpha & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_2 & \lambda_1 \end{pmatrix}, \quad \alpha = 0.67680,$$

$$\lambda_1 = 0.10447, \quad \lambda_2 = 0.25039 \quad \text{and} \quad \lambda_3 = 0.29028.$$

$$\zeta_6^* = \begin{pmatrix} -1 & -\alpha & -\beta & \beta & \alpha & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_3 & \lambda_2 & \lambda_1 \end{pmatrix}, \quad \alpha = 0.78863, \quad \beta = 0.29129,$$

$$\lambda_1 = 0.07992, \quad \lambda_2 = 0.18749 \quad \text{and} \quad \lambda_3 = 0.23259.$$

$$\zeta_7^* = \begin{pmatrix} -1 & -\alpha & -\beta & 0 & \beta & \alpha & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_3 & \lambda_2 & \lambda_1 \end{pmatrix}, \quad \alpha = 0.84267, \quad \beta = 0.43195,$$

$$\lambda_1 = 0.06685, \quad \lambda_2 = 0.15663, \quad \lambda_3 = 0.23830 \quad \text{and} \quad \lambda_4 = 0.07644.$$

## 6. Optimal designs based on MSE

We note here that equivalence theory is based on the assumption that the model is valid. Under this assumption we try to get a good estimator of  $\beta$ , or of linear combinations of  $\beta$ . When the regression model has not been specified correctly, can we still get a good design?

Let us return to the linear model (1.1). Consider the case when the model has not been specified correctly and that the specification uses only the first  $s$  ( $< k$ ) parameters as

$$\text{the valid model: } y = f'_0(x)\beta_0 + f'_1(x)\beta_1 + e, \tag{6.1}$$

$$\text{the fitted model: } y = f'_0(x)\beta_0 + e,$$

where  $f(x)' = (f'_0(x), f'_1(x))$ ,  $f_0(x): s \times 1$ ,  $\beta' = (\beta'_0, \beta'_1)$ ,  $\beta_0: s \times 1$ .

When  $n$  independent observations  $y_1, \dots, y_n$  based on  $n$  explanatory variables  $x_1, \dots, x_n$  are given, let  $X = [X_0, X_1] = [f(x_1), \dots, f(x_n)]'$  be of size  $n \times k$ ,  $X_0$  of size

$n \times s$  and  $\mathbf{y} = (y_1, \dots, y_n)'$ . Using the ordinary least squares method to the fitted model (6.1), we predict the response variable at a point  $z \in \mathcal{X}$  by  $\hat{y}(z) = \mathbf{f}'_0(z)(X'_0 X_0)^{-1} X'_0 \mathbf{y}$  and its mean squared error is given by

$$\begin{aligned} \text{MSE}(z) &= E[\hat{y}(z) - \mathbf{f}'(z)\boldsymbol{\beta}]^2 \\ &= \text{var}[\hat{y}(z)] + [E\hat{y}(z) - \mathbf{f}'(z)\boldsymbol{\beta}]^2 \\ &= \frac{\sigma^2}{n} \mathbf{f}'_0(z) M_{00}^{-1} \mathbf{f}_0(z) \\ &\quad + \boldsymbol{\beta}'_1 [M_{10} M_{00}^{-1}, -I_r] \mathbf{f}(z) \mathbf{f}'(z) [M_{10} M_{00}^{-1}, -I_r]' \boldsymbol{\beta}_1, \end{aligned} \quad (6.2)$$

where  $M$  is an information matrix divided by  $n$  and is partitioned as

$$M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix} = \frac{1}{n} X'X : k \times k, \quad M_{00} : s \times s \quad \text{and} \quad r = k - s.$$

Consider the case that we are interested in minimizing the average of MSE in some sense.

Let  $\mu$  be a given probability measure on the design space  $\mathcal{X}$ .  $\mu$  is used for averaging MSE and we call it an *averaging probability measure*. Let the information matrix of  $\mu$  be

$$L = \begin{pmatrix} L_{00} & L_{01} \\ L_{10} & L_{11} \end{pmatrix} = \int_{\mathcal{X}} \mathbf{f}(z) \mathbf{f}'(z) d\mu(z) : k \times k \quad \text{and} \quad L_{00} : s \times s. \quad (6.3)$$

Obviously  $L$  is a positive semi-definite matrix. Assume  $L_{00}$  is positive definite. (Note that  $L_{00} : s \times s$  corresponds to  $L : k \times k$  used in (5.1).) Integrating the MSE of (6.2) by the averaging measure  $\mu$ , we get the integrated MSE (IMSE):

$$\begin{aligned} \text{IMSE} &= \int_{\mathcal{X}} \text{MSE}(z) d\mu(z) \\ &= \frac{\sigma^2}{n} \text{tr}(M_{00}^{-1} L_{00}) + \boldsymbol{\beta}'_1 [M_{10} M_{00}^{-1}, -I_r] L [M_{10} M_{00}^{-1}, -I_r]' \boldsymbol{\beta}_1 \\ &= -\Psi_0(M) - \Psi_1(M) + \boldsymbol{\beta}'_1 (L_{11} - L_{10} L_{00}^{-1} L_{01}) \boldsymbol{\beta}_1, \end{aligned} \quad (6.4)$$

where

$$\Psi_0(M) = -\frac{\sigma^2}{n} \text{tr}(M_{00}^{-1} L_{00}), \quad (6.5)$$

$$\Psi_1(M) = -\boldsymbol{\beta}'_1 (M_{10} M_{00}^{-1} - L_{10} L_{00}^{-1}) L_{00} (M_{00}^{-1} M_{01} - L_{00}^{-1} L_{01}) \boldsymbol{\beta}_1. \quad (6.6)$$

Note that the last term of (6.4) does not depend on the choice of the information matrix.

We have derived the integrated MSE of (6.4) when the design consists of  $n$  points. Note that the IMSE is a function of designs through its information matrix.

Hence the IMSE can be treated as a function of information matrices of normalized designs.

We partition information matrices as

$$M(\xi) = M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix} = \int_{\mathcal{X}} f(x)f(x)' d\xi(x) : k \times k \quad \text{and} \quad M_{00} : s \times s$$

for probability measures  $\xi$  on the design space  $\mathcal{X}$ . Recall that  $\mathcal{M}$  of (1.6) is a collection of matrices  $M$ , and let  $\mathcal{M}_0$  be a set of  $M$  whose submatrices  $M_{00}$  are positive definite. Then  $\mathcal{M}$  and  $\mathcal{M}_0$  are convex sets in the  $k(k+1)/2$ -dimensional Euclidean space. Thus our problem is to find a design  $\xi$  on  $\mathcal{X}$  or an information matrix  $M = M(\xi) \in \mathcal{M}_0$  maximizing

$$\Psi(M) = \Psi_0(M) + \Psi_1(M), \quad (6.7)$$

where  $\Psi_0(M)$  and  $\Psi_1(M)$  are given in (6.5) and (6.6), respectively. We say that  $M^* = M(\eta^*)$  is a  $\Psi$ -optimal information matrix or  $\eta^*$  is a  $\Psi$ -optimal design if  $\Psi(M)$  of (6.7) takes its minimum value at  $M^* = M(\eta^*)$ .

The function  $\Psi_0(M)$  is well known as a linear criterion function. When  $L_{00} = cc'$ ,  $c'M_{00}^{-1}c$  is known as the *c-optimal criterion function*, see, e.g., [17, p. 13]. Also the *A-optimal criterion function*,  $\text{tr} M_{00}^{-1}$ , is given when  $L_{00} = I_s$ .

The Fréchet derivative of  $\Psi$  at  $M \in \mathcal{M}^+$  in the direction of  $N \in \mathcal{M}$  is expressible, after some algebra, as

$$F_{\Psi}(M, N) = F_{\Psi_0}(M, N) + F_{\Psi_1}(M, N),$$

where  $N$  is partitioned in the same way as  $M$  and

$$F_{\Psi_0}(M, N) = \frac{\sigma^2}{n} \{ \text{tr}(M_{00}^{-1} L_{00} M_{00}^{-1} N_{00}) - \text{tr}(M_{00}^{-1} L_{00}) \}, \quad (6.8)$$

$$F_{\Psi_1}(M, N) = 2\beta_1'(M_{10} M_{00}^{-1} N_{00} - N_{10}) M_{00}^{-1} L_{00} (M_{00}^{-1} M_{01} - L_{00}^{-1} L_{01}) \beta_1. \quad (6.9)$$

The function  $\Psi_0(M)$  of (6.5) is convex and differentiable on  $\mathcal{M}_0$ . Hence the equivalence theorem is valid for  $\Psi_0(M)$ . Though the criterion function  $\Psi_1$  of (6.6) is differentiable, it is not concave when  $\beta_2 \neq \mathbf{0}$ . Furthermore,  $\Psi_1$  takes its minimum value zero when  $M_{10} M_{00}^{-1} = L_{10} L_{00}^{-1}$  i.e. when the design equals the averaging probability measure. Thus the criterion function  $\Psi$  is differentiable in the Fréchet sense at any point in  $\mathcal{M}_0$ . However, the general equivalence theory cannot simply be applied since the function  $\Psi$  on  $\mathcal{M}$  is neither convex nor concave.

**Theorem 6.1** (Nishii [10]). *A necessary condition for  $M^* = M(\xi^*) \in \mathcal{M}$  to maximize  $\Psi$  on  $\mathcal{M}$  is that*

$$F_{\Psi}(M^*, f(x)f(x)') \leq 0 \quad \text{for any } x \text{ in } \mathcal{X}.$$

Furthermore, all the support points  $x^*$  of  $\xi^*$  satisfy

$$F_{\Psi}(M^*, f(x^*)f(x^*)') = 0.$$

The converse of Theorem 6.1 is not true in general since  $\Psi$  is not convex. For the same reason the  $\Psi$ -optimal information matrix may not be unique. Note that there exist more than two designs having the same information matrix in many cases.

$\xi$  is defined as a *minimax design* if the information matrix  $M(\xi)$  minimizes  $\sup_{\beta \in \mathbb{R}^k} \Psi(N)$  for  $N \in \mathcal{M}^+$ . This design is also called *all-bias design*. The following theorem has been established.

**Theorem 6.2** (Nishii [10]). (i) *The averaging probability measure  $\mu$ , which is used for averaging MSE, is a minimax design.*

(ii) *If  $(\sqrt{n}/\sigma) \|\beta_1\|$  is sufficiently small, the  $\Psi_0$ -optimal design still remains  $\Psi$ -optimal.*

## 7. Example of designs minimizing MSE

Consider the single polynomial regression on the interval  $[-1, 1]$  expressed as

$$\text{the valid model: } y = \beta_0 + \beta_1 x + \beta_2 x^2 + e \quad (\beta_2 \neq 0),$$

$$\text{the fitted model: } y = \beta_0 + \beta_1 x + e,$$

where  $x \in [-1, 1]$ . We use a symmetric averaging probability measure  $\mu$  on  $[-1, 1]$ . Let

$$M(\mu) = \left( \int_{-1}^1 x^{i+j-2} d\mu(x) \right) = L = \begin{pmatrix} 1 & 0 & \alpha \\ 0 & \alpha & 0 \\ \alpha & 0 & \gamma \end{pmatrix}, \quad L_{00} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix},$$

where  $\alpha > 0$  is a second moment of the averaging probability measure  $\mu$ . For example, when  $\mu$  is a uniform distribution on  $[-1, 1]$ ,  $\alpha$  is given by  $1/3$ . Let  $\xi$  be a design on  $[-1, 1]$ . We denote its first, second and third moments by  $a$ ,  $b$  and  $c$ , respectively. Then

$$M(\xi) = \begin{pmatrix} 1 & a & b \\ a & b & c \\ b & c & d \end{pmatrix}, \quad M_{00} = \begin{pmatrix} 1 & a \\ a & b \end{pmatrix}.$$

Hence the integrated mean squared error based on the design  $\xi$  is

$$\Psi(M(\xi)) = \Psi_0(M(\xi)) + \Psi_1(M(\xi)),$$

where

$$\Psi_0(M(\xi)) = -(\sigma^2/n) \{1 + (a^2 + \alpha)/(b - a^2)\},$$

$$\Psi_1(M(\xi)) = -\beta_2^2 [\{b - \alpha - a(c - ab)/(b - a^2)\}^2 + \alpha \{(c - ab)/(b - a^2)\}^2].$$

Note that the fourth moments of  $\mu$  and of  $\xi$  do not appear in the function  $\Psi(M(\xi))$ . Obviously the unique  $\Psi_0$ -optimal design is

$$\begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

and the  $\Psi_1$ -optimal design (minimax design) is

$$\begin{pmatrix} -\sqrt{\alpha} & \sqrt{\alpha} \\ 1/2 & 1/2 \end{pmatrix}.$$

Both designs are symmetric.

**Theorem 7.1** (Nishii [10]). *Let  $\alpha$  be the second moment of the symmetric averaging probability measure  $\mu$ . Let the  $\Psi$ -optimal design be  $\xi^*$ .*

*Case 1: If  $\alpha = 1$ , i.e.,*

$$\mu = \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix},$$

*then*

$$\xi^* = \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}.$$

*Case 2: If  $0 < \alpha < 1$ ,*

$$\text{when } 0 \leq \frac{n}{\sigma^2} \beta_2^2 \leq \alpha / \{2(1-\alpha)\}; \quad \text{then } \xi^* = \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix},$$

$$\text{when } \frac{n}{\sigma^2} \beta_2^2 > \alpha / \{2(1-\alpha)\}; \quad \text{then } \xi^* = \begin{pmatrix} -\sqrt{t^*} & \sqrt{t^*} \\ 1/2 & 1/2 \end{pmatrix},$$

*where*

$$t^* = \left\{ \frac{c + \sqrt{D}}{2} \right\}^{1/3} + \left\{ \frac{c - \sqrt{D}}{2} \right\}^{1/3} + \frac{\alpha}{3},$$

$$c = \frac{2\alpha^3}{27} + \frac{\alpha}{2\gamma^2}, \quad D = \frac{\alpha^2}{\gamma^2} \left( \frac{2\alpha^3}{27} + \frac{1}{4\gamma^2} \right) \quad \text{and} \quad \gamma = (\sqrt{n}/\sigma)\beta_2.$$

**Outline of the proof.** First the optimal design  $\xi^*$  is shown to be symmetric, i.e., first and third moments are zero. Let  $M^* = M(\xi^*)$  ( $\Psi$ -optimal information matrix) be

$$M^* = \begin{pmatrix} 1 & 0 & b \\ 0 & b & 0 \\ b & 0 & d \end{pmatrix}, \quad 0 < b \leq 1.$$

Theorem 6.1, (6.8) and (6.9) yield

$$F_{\Psi}(M^*, \mathbf{x}\mathbf{x}') = (x^2 - b)(\alpha + 2\alpha\gamma^2b^2 - 2\alpha^2b^3)/b^2 \leq 0 \quad (7.1)$$

for any  $x \in [-1, 1]$  where  $\mathbf{x} = (1, x, x^2)'$ . Let  $h(b) = \alpha + 2\alpha\gamma^2b^2 - 2\alpha^2b^3$ . The inequality (7.1) has two solutions. One is

$$b = 1 \quad \text{and} \quad h(1) = \alpha - 2(1 - \alpha)\gamma^2 \geq 0,$$

and the other is

$$0 < b < 1 \quad \text{and} \quad h(b) = 0.$$

In the second case  $h(b) = 0$  has the unique solution between  $2\alpha/3$  and 1 if and only if  $h(1) < 0$ . We note that the  $\Psi$ -optimal design of this case is symmetric.  $\square$

## 8. Related unsolved problems

In this section we present open problems. Consider polynomial regression of the single control variable  $x \in [-1, 1]$

$$\text{the valid model: } y = \sum_{i=0}^{k-1} \beta_i x^i + e,$$

$$\text{the fitted model: } y = \sum_{i=0}^{s-1} \beta_i x^i + e,$$

for  $s < k$ , and let  $L$  be the information matrix of a symmetric averaging measure  $\mu$  on  $[-1, 1]$  partitioned as (6.3).

**Problem 8.1.** Let the probability measure  $\eta^*$  minimize the IMSE of (6.4). Then:

- (a) verify whether  $\eta^*$  is symmetric or not, and
- (b) express  $\eta^*$  in terms of the elements of  $L$  and  $\beta$ .

It is true that  $\eta^*$  coincides with the averaging measure  $\mu$  when  $(\beta_s, \dots, \beta_{k-1})$  is far from  $(0, \dots, 0)$ . Also the IMSE is invariant under the transformation  $x \rightarrow -x$  because  $\mu$  is assumed to be symmetric. Unfortunately the integrated MSE is neither convex nor concave. Hence (a) should be examined. Note that the result based on the group invariance by Pukelsheim [14] is not applicable. In the previous section, the problem in the case  $k=3$  and  $s=2$  is positively solved.

A generalization into a multiple control variable case like Hoel [4] is also interesting. Let  $(x_1, \dots, x_m)$  be a vector of control variables on the design space

$[-1, 1]^m \equiv [-1, 1] \times \cdots \times [-1, 1]$ . When  $s < k$ , consider polynomial regression with  $m$  control variables

$$\text{the valid model: } y = \sum_{\delta_1=0}^{k_1} \cdots \sum_{\delta_m=0}^{k_m} \beta_{\delta_1 \dots \delta_m} x_1^{\delta_1} \cdots x_m^{\delta_m} + e, \quad (8.1)$$

$$\text{the fitted model: } y = \sum_{\delta_1=0}^{s_1} \cdots \sum_{\delta_m=0}^{s_m} \beta_{\delta_1 \dots \delta_m} x_1^{\delta_1} \cdots x_m^{\delta_m} + e, \quad (8.2)$$

where  $s_1 \leq k_1, \dots, s_m \leq k_m$ . In this multiple variable case, the probability measure  $\eta$  is defined to be *symmetric* if the measures  $\eta(\pm x_1, \dots, \pm x_m)$  coincide with  $\eta(x_1, \dots, x_m)$ .

**Problem 8.2.** Let  $\mu_m$  be a symmetric probability measure on  $[-1, 1]^m$ , let  $L_m$  be the  $k^m \times k^m$  information matrix of  $\mu_m$ , and further let the probability measure  $\eta_m^*$  minimize the IMSE of (6.4). Then:

- (a) verify whether  $\eta_m^*$  is symmetric or not, and
- (b) express  $\eta_m^*$  in terms of the elements of  $L_m$  and  $\beta_m \equiv (\beta_{0 \dots 0}, \dots, \beta_{k-1 \dots k-1})'$ .

We note here that the approach used in [16] is not applicable because the IMSE does not fulfill the requirement for the derivative. Thus it is not sure whether  $\eta_m^*$  can be expressed as the product measure of optimal designs with single control variable.

Another generalization with  $m$  control variables is given:

$$\text{the valid model: } y = \sum_k^* \beta_{\delta_1 \dots \delta_m} x_1^{\delta_1} \times \cdots \times x_m^{\delta_m} + e, \quad (8.3)$$

$$\text{the fitted model: } y = \sum_s^* \beta_{\delta_1 \dots \delta_m} x_1^{\delta_1} \times \cdots \times x_m^{\delta_m} + e, \quad (8.4)$$

where  $\sum_k^*$  extends over all nonnegative integers  $\delta_1, \dots, \delta_m$  such that  $\delta_1 + \cdots + \delta_m \leq k$ , and  $\sum_s^*$  is similarly defined. Replacing (8.1) and (8.2) by (8.3) and (8.4) in the previous problem, we obtain another problem.

## Acknowledgement

The author would like to express his thanks to the referees for their valuable comments and careful reading of the early draft. Also thanks are due to the language editor for improving the expression of the draft.

## References

- [1] A.C. Atkinson, Recent developments in the methods of optimum and related experimental designs, *Internat. Statist. Rev.* 56 (1988) 99–115.
- [2] V.V. Fedorov, *Theory of Optimal Experiments* (translated and edited by W.J. Studden and E.M. Klimko) (Academic Press, New York, 1972).



- [3] N. Gaffke, Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression, *Ann. Statist.* 15 (1987) 942–957.
- [4] P.G. Hoel, Optimum designs for polynomial extrapolation, *Ann. Math. Statist.* 36 (1965) 1483–1493.
- [5] S. Karlin and W.J. Studden, *Techebycheff Systems: With Applications in Analysis and Statistics* (Interscience, New York, 1966).
- [6] J. Kiefer and J. Wolfowitz, The equivalence of two extremum problems, *Canad. J. Math.* 12 (1960) 363–366.
- [7] J. Kiefer, General equivalence theory for optimum designs (approximate theory), *Ann. Statist.* 2 (1974) 849–879.
- [8] J.N.S. Matthews, Recent developments in crossover designs, *Internat. Statist. Rev.* 56 (1988) 117–127.
- [9] W. Näther and V. Reinsch,  $D_\lambda$ -optimality and Whittle's equivalence theorem, *Math. Oper. Statist., Ser. Statist.* 12 (1981) 307–316.
- [10] R. Nishii, Optimal designs minimizing mean squared error for polynomial regression, *J. Japan Statist. Soc.* 17 (1987) 113–118.
- [11] L.J. Paterson, Some recent work on making incomplete-block designs available as a tool for science, *Internat. Statist. Rev.* 56 (1988) 129–138.
- [12] A. Pázman, *Foundations of Optimum Experimental Design* (Reidel, Dordrecht, 1986).
- [13] F. Pukelsheim, On linear regression designs which maximize information, *J. Statist. Plann. Inference* 4 (1980) 339–364.
- [14] F. Pukelsheim, Information increasing orderings in experimental design theory, *Internat. Statist. Rev.* 55 (1987) 203–219.
- [15] F. Pukelsheim and D.M. Titterington, General differential and Lagrangian theory for optimal experimental designs, *Ann. Statist.* 11 (1983) 1060–1068.
- [16] E. Rafajlowicz and W. Myszka, Optimum experimental design for a regression on a hypercube-generalization of Hoel's result, *Ann. Inst. Statist. Math.* 40 (1988) 821–827.
- [17] S.D. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation* (Chapman and Hall, London, 1980).
- [18] P. Whittle, Some general points in the theory of optimal experimental design, *J. Roy. Statist. Soc. B.* 35 (1971) 123–130.